한국컴퓨터그래픽스학회
Korea Computer Graphics Society Vol. 30, No. 2, P. 21~29

강화학습을 이용한 클라이밍 모션 합성

강경원° 권태수*

한양대학교 일반대학원 컴퓨터소프트웨어학과

ruddnjs925@hanyang.ac.kr, taesoobear@gmail.com

Climbing Motion Synthesis using Reinforcement Learning

Kyungwon Kang° Taesoo Kwon*

Dept. of Computer and Software, Hanyang University

요 약

최근 자연스러운 모션 데이터에 대한 수요가 늘고 있지만, 클라이밍 모션을 정확하게 캡처하는 것은 가려진 부분이 많은 클라이밍 동작의 특성상 쉽지 않다. 또한 벽 구조물의 스캔이나 다양한 암벽 코스 준비 등 필요한 데이터를 수집하는 과정이 쉽지 않다. 본 논문에서는 강화학습을 이용한 클라이밍 모션 합성 방법론을 제안한다. 학습 과정은 두 단계의 난이도로 구성되어 있다. 첫 번째 단계는 매달리기 정책을 학습하는 것이다. 매달리기 정책은 자연스러운 자세로 홀드를 잡는 방법을 학습한다. 이후 추론 단계를 통해 위치, 자세, 잡기 상태를 다양하게 추출한 초기 상태 데이터세트를 만든다. 두 번째 단계에서는 이 초기 상태 데이터세트를 사용해서 실제 클라이밍을 수행하는 태스크를 학습한다. 클라이밍 정책은 자연스러운 자세로 타겟 위치로 이동하는 방법을 학습한다. 실험을 통해 제안하는 방법이 클라이밍 하기 위한 좋은 자세를 효과적으로 탐색할 수 있는 것을 보였다.

Abstract

Although there is an increasing demand for capturing various natural motions, collecting climbing motion data is difficult due to technical complexities, related to obscured markers. Additionally, scanning climbing structures and preparing diverse routes further complicate the collection of necessary data. To tackle this challenge, this paper proposes a climbing motion synthesis using reinforcement learning. The method comprises two learning stages. Firstly, the hanging policy is trained to grasp holds in a natural posture. Once the policy is obtained, it is used to extract the positions of the holds, postures, and gripping states, thus forming a dataset of favorable initial poses. Subsequently, the climbing policy is trained to execute actual climbing maneuvers using this initial state dataset. The climbing policy allows the character to move to the target location using limbs more evenly in a natural posture. Experiments have shown that the proposed method can effectively explore the space of good postures for climbing and use limbs more evenly. Experimental results demonstrate the effectiveness of the proposed method in exploring optimal climbing postures and promoting balanced limb utilization.

키워드: 캐릭터 애니메이션, 물리 시뮬레이션, 강화학습, 클라이밍
Keywords: Character Animation, Physics Simulation, Reinforcement Learning, Climbing

1 서론

자연스러운 움직임을 캡쳐하는 기술은 영상, 게임, VR, 로보틱스 등 다양한 산업에서 꾸준히 수요가 증가하고 있다. 하지만 모션 캡처 기술을 통해 좋은 품질의 모션을 얻기 위해서는 인적, 시간적, 물질적 비용이 많이 필요하다. 컴퓨터 비전 분야와 컴퓨터 그래픽스 분야에서는 인공지능 연구의 발전과 함께, 더 적은 비용으로 고품질의 모션을 획득하기 위한 기술들이 연구되고 있다. 최근에는 단안 카메라를 통해 모션[1]을 획득하는 방법, 적은 수

*corresponding author: Taesoo Kwon / Dept. of Computer and Software, Hanyang University (taesoobear@gmail.com)

Received : 2024.04.08./ Review completed : 1st 2024.05.06. / Accepted : 2024.05.13.
DOI : 10.15701/kcgs.2024.30.2.21
ISSN : 1975-7883(Print)/2383-529X(Online)

의 IMU 센서 장비를 통해 모션을 획득하는 방법[2]이 연구되고 있다. 이러한 연구들은 캡처를 수행하는 장비의 비용이나 제약조건을 줄이는 것에 집중하고 있다. 캡처를 수행하지 않고 모션을 획득할 수 있는 다른 연구로는, 물리 시뮬레이션 환경에서 캐릭터를 컨트롤하여 모션을 합성하는 방법이 있고, 최근 강화학습을 이용한 방법이 활발히 연구되고 있다 [3, 4, 5]. 이러한 방법은 게임이나 영화 프로토타이핑 등 인터액티브 앱에서도 활용할 수 있다는 장점이 있다.

클라이밍(climbing)은 사람이 사지를 이용해서 벽이나 가파른 경사로를 기어오르는 행동이다. 클라이밍 모션을 캡처하기 위해서는 모션 캡처 장비뿐만 아니라, 기어오를 클라이밍 벽 구조물 또한 필요하다. 촬영한 모션을 재사용하기 위해서는 구조물도 같이 촬영해야 한다. 또한, 한 번 제작한 인공 벽 구조물은 수정이 어려운 단점이 있다. 결과적으로, 클라이밍 모션을 촬영하기는 어려우며, 다양한 장면에 사용하기 위해서는 모션을 수정하는 애니메이터의 후처리 노력이 수반된다.

이러한 문제를 해결하기 위해, 본 논문에서는 물리 환경에서 클라이밍 벽 구조물이 주어졌을 때, 강화학습을 이용해서 물리적으로 설득력 있는 클라이밍 모션을 합성할 수 있는 캐릭터 제어 방법을 제안한다. 제안하는 방법의 구성은 두 단계로 이루어져 있다. 첫 번째는 매달리기 정책이다. 매달리기 정책은 벽에 위치한 손잡이, 홀드(hold)를 잘 잡는 방법을 학습한다. 일반적으로 어디에 캐릭터를 위치시켜도 지지할 바닥을 가까이 두고 시작하는 보행 동작과는 달리, 클라이밍의 경우 면적이 적고, 적당한 홀드를 잡거나 밟아야 떨어지지 않고 몸을 지지할 수 있기 때문에, 캐릭터의 초기 자세를 결정하기 어렵다. 매달리기 정책은 클라이밍 벽 주변 무작위 위치에서 초기 상태가 결정되었을 때, 가능한 자연스러운 자세로 홀드를 많이 붙잡고 있는 것을 목표로 학습한다. 매달리기 정책의 학습이 완료되면, 추론 단계에서 특정 개수 이상 홀드를 동시에 붙잡고 있는 자세를 선별하여, 위치, 자세, 잡기 여부를 데이터 세트(data set)로 저장한다. 두 번째 단계는 클라이밍 정책을 학습하는 단계이다. 매달리기 정책을 통해 획득한 데이터 세트를 이용해 초기 상태 위치, 자세, 잡기 여부를 결정하고, 지정하는 위치까지 캐릭터가 클라이밍 하는 제어 방법을 학습한다. 두 단계 모두 자연스러운 모션 표현을 위해 기존연구 AMP(Adversarial Motion Prior)[4]의 방법론을 활용했다. 학습 결과, 주어진 클라이밍 벽 구조물에서 모든 사지를 이용하여 자연스럽게 지정하는 위치까지 이동하는 결과를 보였다. 본 논문에서 제안하는 방법을 정리하면 다음과 같다.

- 물리 시뮬레이션 환경에서 클라이밍 벽 구조물이 주어졌을 때, 자연스러운 클라이밍 모션 합성을 위해 캐릭터를 제어하는 2단계 학습 방법론을 제안했다.

- 매달리기 정책을 통해서 얻은 초기 상태 데이터세트를 활용하여, 더 자연스럽고 성공률 높은 클라이밍 제어를 보였다.

2 관련 연구

캐릭터 애니메이션 분야에서는 더욱 다양하고 자연스러운 움직임을 합성하는 방법이 연구되고 있다. 연구는 크게 두 분야로 분류할 수 있다. 운동학(kinematic)적 방법과 물리 기반(physics-based)방법이다. 본 장에서는 물리 기반 캐릭터 컨트롤 관련 연구와 클라이밍 모션 합성 관련 연구들을 소개한다.

2.1 물리 기반 캐릭터 컨트롤

물리 기반 캐릭터 컨트롤 분야는, 물리 시뮬레이션 환경에서 토크를 이용해서 캐릭터를 직접 제어하는 방법을 연구하는 분야이다. 최근, 강화학습의 발전과 함께 자연스러운 모션을 갖도록 캐릭터를 제어하는 방법론들이 활발히 연구되고 있다. Peng 등은 강화학습과 모션 캡처 클립을 이용해서 타겟 모션을 모방하는 방법을 제안했다 [3]. 하지만, 특정한 모션을 추적하는 방법으로 새로운 모션을 생성하려면 타겟 모션이 준비되어야 하고, 재학습이 필요한 단점이 있다. Peng 등은 GAIL[6]방법론을 물리 기반 캐릭터 컨트롤에 적용했다 [4]. 모션 캡처 클립과 분류기(discriminator)를 이용해서 분류기가 캐릭터의 움직임을 가짜로 판별하도록 학습하고, 강화학습 보상 함수를 통해 캐릭터가 점점 모션 클립과 비슷한 모션을 취하도록 하는 AMP(Adversarial Motion Prior)를 제안하였다 [4]. 이 방법론은 모방할 모션 캡처 클립을 구성하는 데 필요한 노력이 적은 것이 큰 장점이다. 하지만, 다른 모션 스타일을 사용하고 싶으면 매번 처음부터 다시 학습해야 하는 한계가 있었다. Peng 등은 이를 해결하기 위해서 ASE라고 불리는, 재사용 가능한 저수준(low-level)컨트롤러와 특정 작업을 위한 고수준(high-level)컨트롤러를 이용하는 효율적인 학습 방법론을 제안했다 [5]. 이후의 최신 연구들은 저수준 컨트롤러에서 특정 컨디션에 대한 모션을 생성할 수 있게 하는 방법 [7, 8], GAN 방식의 한계를 극복하기 위해 잠재 공간의 표현을 달리한 방법 [9] 등이 있다. 본 연구에서는 AMP[4] 방법론을 활용했다. 클라이밍 작업은 범용 저수준 컨트롤러 제작이 어렵기 때문에, 재사용은 고려하지 않았다.

2.2 클라이밍 모션 합성

클라이밍 모션을 합성하는 연구들은 과거부터 이어져오고 있다. Naderi 등은 물리 환경에서 저수준 최적화 기법과 고수준 경로 계획법을 사용한 클라이밍 움직임 계획법을 제안했다[10]. 하지만, 새 홀드를 잡기 위해 한 번에 움직일 수 있는 사지의 수가 2개라는 한계가 있다. 그 후로 Naderi 등은 인공지능 신경망을 이용해서 이전의 연구보다 더 좋은 성공률과 모든 사지를 이용해 점프 모션이 가능한 것을 보였다. [11] 하지만, 움직임 최적화에 걸리는 시간 때문에, 실시간으로 사용하기 어려운 한계가 있다. Naderi 등은 클라이밍 모션 합성에 강화학습을 적용하였다[12]. 이 방법은 캐릭터의 자세에서 신장, 홀드와의 거리 등의 요소로 타겟 홀드를 추출하고, 선택한 홀드를 잡도록 제어하였다. 하지

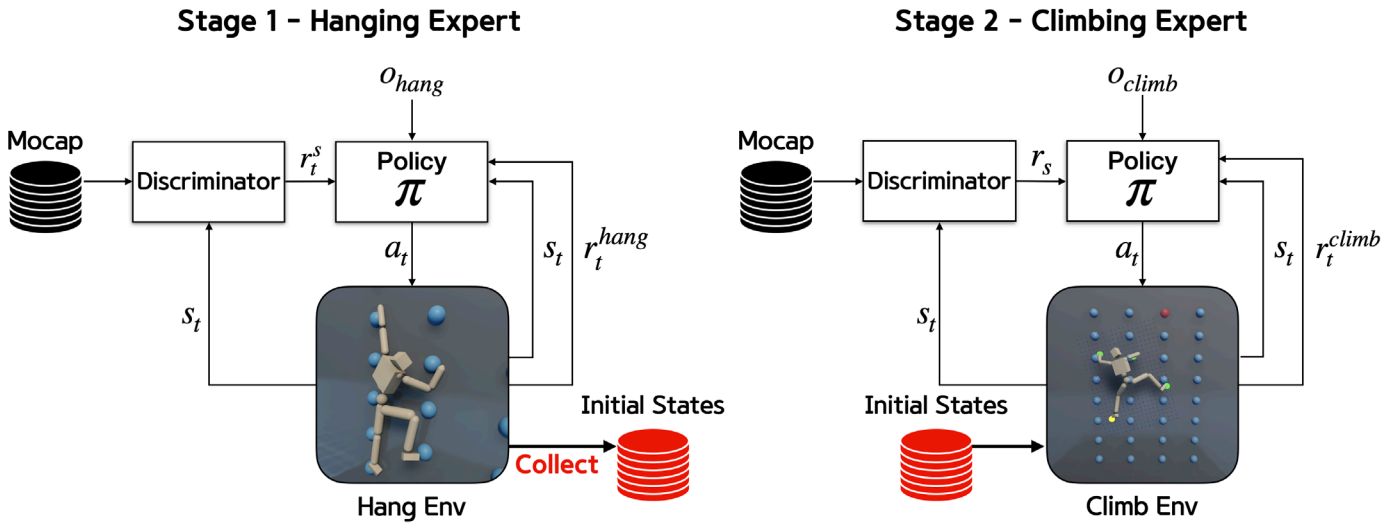


Figure 1: The system comprises two stages of training. During the first stage, it learns to hang in natural poses and collects the initial states using the acquired policy. In the second stage, it learns a climbing policy using the collected initial state dataset.

만, 한 번에 1개 또는 2개의 사지만 움직일 수 있고, 동작이 자연스럽지 못한 한계가 있다. 본 논문에서는 자연스러운 클라이밍 모션의 합성을 위해 AMP[4]와 초기 상태 데이터세트를 활용한다. 강화학습 정책이 깊이 정보를 통해 주변 환경을 관측하고, 잡을 홀드를 스스로 찾아 움직이게 한다.

3 시뮬레이션 환경

Figure 1은 제안하는 방법의 전체적인 흐름도이다. 본 장에서는 사용된 물리 시뮬레이션 환경을 소개하고, 4장에서 강화학습 학습 방법을 설명한다.

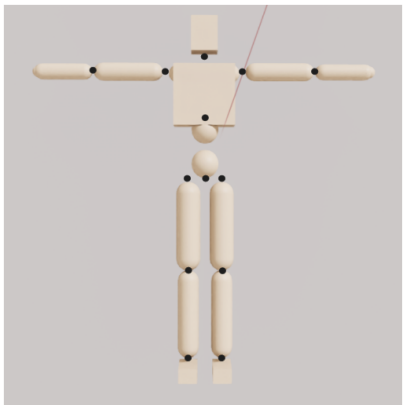
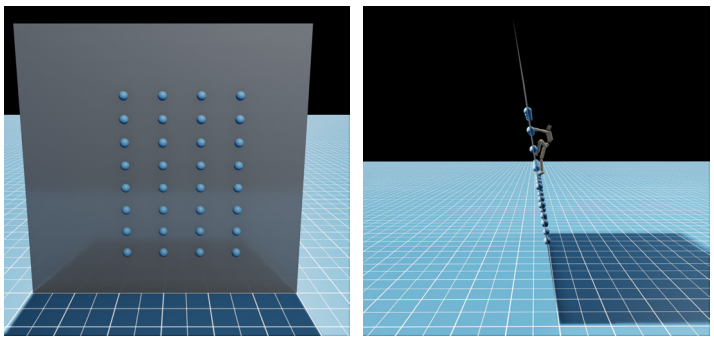


Figure 2: Our character model. The black dots represent the character's joint positions.

3.1 캐릭터 모델과 클라이밍 벽

Figure 2는 실험에 사용된 캐릭터이다. 루트 조인트를 포함한 조인트는 13개이고, 자유도(Degrees of Freedom)는 28이다. 신장은



(a) (b)

Figure 3: The climbing wall used in the experiment. (a) Frontal view of the climbing wall, (b) Side view of the climbing wall

1.75m, 무게는 47kg이다. 각 관절은 PD 제어기를 이용해 제어했다. Figure 3은 실험에 사용된 클라이밍 벽이다. 벽의 경사는 약 80°이다. 홀드는 총 16개를 사용했으며, 반지름이 10cm인 구체 형태이다. 홀드의 위치는 4x8 격자에 가로 80cm, 세로 50cm의 동일한 간격을 갖게 설정했다. 가장 높은 위치에 있는 홀드의 높이는 4.5m이다. 실제로 캐릭터 모델이 홀드를 붙잡려면, 손 조인트를 구현해야 하고, 이는 제어할 자유도가 크게 늘어나기 때문에 제어하기 어려워진다. 문제의 난이도를 낮추기 위해, 캐릭터 모델이 홀드를 잡고 밟는 것을 물리 환경에 제약조건을 추가하는 방법으로 구현한다. 정책 네트워크로부터 각 사지의 잡기 신호가 출력되고, 캐릭터 사지의 종단 위치가 홀드의 중심과 일정 거리 d 안에 있으면, 해당 바디 위치를 고정하는 제약조건을 사용한다. 실험에서 상체는 손목, 하체는 발가락을 종단 위치로 사용하였다.

3.2 매달리기 정책

Figure 1의 stage1은 매달리기 정책의 흐름도를 나타낸다. 이 단계는 클라이밍 정책 (Figure 1 stage2)의 더 자연스러운 모션과 높은 성공률 학습을 위해 필요한 단계이다. 클라이밍 정책을 학습할 때, 무작위 위치에서 떨어뜨리는 방식으로 상태 초기화를 수행하면 다리는 주도적으로 잡기를 수행하지 않고 두 팔만을 이용해서 클라이밍을 수행하는 경향을 보이는 것을 실험을 통해 확인했다. 원인은 클라이밍을 수행할 때, 팔과 다리를 자연스럽게 동시에 잡는 것을 경험하기 어렵기 때문이다. 이러한 방식으로 학습되면, 다리의 잡기는 단지 팔의 부담을 잠시 덜기 위해 사용할 뿐, 능동적으로 다리를 이용하지 않는다. 이를 해결하기 위해, 초기 상태에서 팔과 다리를 동시에 잡기 수행하는 자세를 경험하게 함으로써, 해당 자세가 자연스러운 클라이밍에 도움이 된다는 것을 모델에게 알리는 방법을 사용한다. 매달리기 정책은 캐릭터가 클라이밍 벽 주변 무작위 위치에 모션 캡처 데이터세트에서 무작위로 추출한 자세로 떨어졌을 때, 홀드를 최대한 많이 잡는 방법을 학습한다. 매달리기 정책 학습이 완료되면 추론 단계에서 에피소드가 종료될 때, 모든 사지가 잡기에 성공한 경우, 해당 위치, 자세, 잡기 여부를 추출한다. 이를 클라이밍 정책의 초기 상태 데이터세트로 사용함으로써, 캐릭터는 더 다양한 자세와 잡기 상태를 경험할 수 있고, 이는 자연스러운 클라이밍 모션 학습을 수행하게 할 수 있다. 실험에서 d 값은 15cm로 설정했다. 사지의 위치가 홀드의 표면과 5cm 거리 안에 존재할 때 잡을 수 있다.

3.3 클라이밍 정책

클라이밍 정책은 클라이밍 벽에 목표 위치가 주어졌을 때, 해당 위치로 캐릭터의 루트를 위치시키는 작업을 학습한다. d 값은 20cm로 설정했다. 전 단계에서 얻은 초기 상태 데이터세트를 이용해서 초기 상태를 설정한다. 실험을 통해 확인한 결과, 초기 상태 데이터세트만을 이용해서 초기 상태를 결정할 경우, 시작한 상태 그대로 잡기를 수행한 뒤, 다른 홀드를 탐색하지 못하고 같은 위치에 있는 결과를 확인했다. 이를 해결하기 위해서 무작위 위치 초기화 방법과 비율을 적절히 나눠 실험에 적용한다. 0.8의 확률로 데이터세트, 0.2의 확률로 무작위 위치로 설정한다.

4 클라이밍 제어 정책 학습

본 장에서는 매달리기, 클라이밍 정책의 강화학습을 위한 상태 공간, 행동 공간, 보상 함수를 정의한다.

4.1 상태 공간

환경에 사용되는 캐릭터 상태의 집합 $s_t \in R^{105}$은 다음과 같이 정의한다. t는 현재 시뮬레이션 타임스텝을, $t-1$은 이전 타임스텝을 의미한다.

$$s_t = \{d_t^{root},\ q_t^r,\ v_t,\ \omega_t,\ q_t,\ \dot{q}_t,\ k_t\}$$

$d_t^{root} \in R^1$는 각 캐릭터 환경 시작 위치에 상대적인 루트의 x 축 좌표이다. 루트의 깊이 위치를 알기 위해 사용된다. $q_t^r \in R^6$, $v_t \in R^3$, $\omega_t \in R^3$는 각각 루트 조인트의 방향, 속도, 각속도를 의미한다. $q_t \in R^{52}$는 루트 조인트를 제외한 조인트의 각도이고, $\dot{q}_t \in R^{28}$은 그 조인트들의 속도를 의미한다. $k_t \in R^{12}$는 잡기를 수행하는 주요 바디의 위치를 의미한다. 루트 방향, 속도, 가속도, 주요 바디는 캐릭터의 로컬 좌표계에 표현된다. 로컬 좌표계는 yz평면상에서의 루트 조인트의 헤딩 방향으로 정의된다. AMP[4]와 같이, 루트 조인트의 방향과 3축 볼 조인트의 방향은 탄젠트-노멀로 표현된다.

매달리기 정책을 학습하기 위한 $o_{hang} \in R^{483}$은 다음과 같다.

$$o_{hang} = \{d_t,\ g_{t-1},\ f_t\}$$

$d_t \in R^{475}$는 클라이밍 벽의 깊이를 나타낸다. 1.9m x 2.5m 크기의 격자 형태, 10cm의 정확도를 갖는다. 깊이 정보를 통해서 클라이밍 벽에서 벽과 홀드의 위치를 구별할 수 있다. $g_{t-1} \in R^4$은 행동 값으로 나온 각 주요 바디가 잡을지 놓을지에 대한 신호이다. $f_t \in R^4$는 실제로 각 주요 바디가 홀드를 잡고 있는지 놓고 있는지를 나타낸다.

클라이밍 정책을 학습하기 위한 $o_{climb} \in R^{489}$은 다음과 같다.

$$o_{climb} = \{o_{hang},\ GT_t,\ x^*\}$$

$GT_t \in R^4$는 주요 바디들 각각의 잡기 또는 놓기 상태의 지속 시간을 나타낸다. $x^* \in R^2$는 로컬 좌표계에 상대적으로 표현된 yz 평면 위 타겟의 위치이다.

4.2 행동 공간

매달리기, 클라이밍 정책은 같은 행동 공간 $a \in R^{32}$를 가진다.

$$a = \{z,\ g\}$$

$z \in R^{28}$은 PD 컨트롤을 위한 각 조인트의 타겟 각도이다. 잡을지, 놓을지에 대한 신호 g를 정책이 결정한다.

4.3 조기 종료

학습에서 에피소드 종료 조건은, 시뮬레이션 시간이 지정한 타임스텝 t를 넘기면 종료된다. 실험에서는 t=300이 사용되었다. 학습을 돕기 위해, 캐릭터가 회복 불가능한 상태가 되면 실패로 판단하고 에피소드를 종료하도록 조기 종료를 적용한다. 회복 불가능한 상황은 매달리기, 클라이밍 행동을 수행하기 불가능한 상태를 의미한다. 조기 종료 조건은 네 가지를 사용한다. 첫 번째, 글로

벌 프레임 기준 루트의 높이가 0.9m보다 낮으면 땅에 닿았다고 판단하여 종료시킨다. 두 번째, 머리 바디에 접촉이 있을 때 종료시킨다. 세 번째, 루트의 수직 속도가 -5m/s 이하면 떨어진다고 판단하고 종료시킨다. 네 번째, 루트에서 가슴까지 방향벡터와 글로벌 업 벡터 차이가 90° 이상이면 종료시킨다.

4.4 보상 함수

벽 모양에 대응해 자연스러운 자세를 가지며 동시에 최대한 많은 홀드를 잡도록 하는 매달리기 정책의 보상 함수는 다음과 같이 정의한다. r_{num_grab}는 모델이 홀드를 동시에 최대한 많이 잡도록 유도하는 보상 함수이다. K는 잡기 여부를 판단할 바디의 수이며, $K = 4$이다. r_{close}는 홀드를 잡을 때, 최대한 벽에 밀착하도록 유도하는 보상 함수이다. x^{wall}은 현재 루트 위치에서 벽에 사영된 좌표의 글로벌 포지션이다. r_{chest}는 캐릭터의 상체 헤딩 방향 d_{chest}이 벽의 기울기에 맞게 유도하는 보상 함수이다. 벽의 법선 벡터 n_{wall}와 캐릭터 상체 방향과 코사인 유사도 $\langle , \rangle$를 이용해서 차이를 작게 유도한다. r_{up}은 상체가 눕는 것을 방지하는 보상 함수로, 글로벌 업 벡터 d_{up}와 루트에서 가슴까지 방향 벡터 d_{r2c}의 차이를 줄인다. 각 보상 함수를 수식으로 표현하면 다음과 같다.

$$r_{num\_grab} = \sum_{k=1}^{K} w_{num\_grab} \cdot f_k, \tag{1}$$

$$r_{close} = exp\left[-0.2 \left\| x^{wall} - x^{root} \right\|^2\right], \tag{2}$$

$$r_{chest} = exp\left[-3 \left\| 1 - \langle -n_{wall}, d_{chest} \rangle \right\|^2\right], \tag{3}$$

$$r_{up} = exp\left[-3 \left\| 1 - \langle d_{up}, d_{r2c} \rangle \right\|^2\right]. \tag{4}$$

매달리기 정책의 최종 보상 함수 수식은 다음과 같다.

$$r_{hang} = r_{num\_grab} + r_{close} + r_{chest} \times r_{up}. \tag{5}$$

실험에서 팔 바디의 경우 $w_{num\_grab} = 0.25$, 다리 바디의 경우 $w_{num_grab} = 0.5$로 설정했다.

클라이밍 정책의 보상 함수는 다음과 같이 정의한다. r_{pos}는 타겟 x^*과 루트의 거리 차이를 줄이는 보상 함수이다. r_{mov}는 캐릭터의 움직이는 방향 d_{mov}이 루트에서 타겟까지의 방향 d^*과 일치하도록 유도하는 보상 함수이다. r_{speed}는 루트가 타겟 방향으로 타겟 속력 s^*을 갖도록 유도하는 보상 함수이다. 실험에서 타겟 속력은 $s^* = 1$m/s로 고정했다.

$$r_{pos} = exp\left[-0.3 \left\| x^* - x^{root} \right\|^2\right], \tag{6}$$

$$r_{mov} = d^* \cdot d_{mov}, \tag{7}$$

$$r_{speed} = exp\left[-4 \left(s^* - d^* \cdot v\right)^2\right]. \tag{8}$$

r_{grab} 보상 함수는 캐릭터가 동시에 최대한 많은 홀드를 잡도록 유도하며, 새로운 홀드를 탐색하는 용도의 함수이다. (1) 수식과는 달리, 클라이밍 정책에서의 r_{num_grab}은 r_{grab}의 구성 요소로 사용된다. 새 잡기가 없을 때는 3개를 잡게 유도하고, 새 잡기가 있을 때 4개를 잡게 유도해서 안정성 있는 자세로 클라이밍을 수행하게 한다. 실험을 통해 확인한 결과, 3개만을 잡도록 유도시키면 두 다리 중 하나는 안 쓰도록 학습하고, 4개만을 잡도록 유도시키면 전혀 움직이지 않는 것을 확인했다. r_{time}은 모든 잡기 또는 놓기 유지시간의 평균을 낮추는 보상 함수로, 사지 골고루 잡기 상태를 변하게 한다. r_{new}은 새로운 잡기를 계속 유도하는 보상 함수다. r_{new}만을 사용하면 같은 위치에서 계속 잡았다 놨다를 반복하기 때문에, 이전에 잡은 홀드와 동일한 홀드를 잡으면 페널티를 부여하는 r_{same}을 같이 사용한다. 수식은 다음과 같다.

$$r_{num\_grab} = \begin{cases} exp\left[-\left|4 - \sum_{k=1}^{K} f_k\right|\right] & \text{if } num\_new\_grab > 0 \\ exp\left[-\left|3 - \sum_{k=1}^{K} f_k\right|\right] & \text{otherwise} \end{cases}, \tag{9}$$

$$r_{time} = exp\left[-0.8 \frac{1}{K} \sum_{k=1}^{K} GT_k\right], \tag{10}$$

$$r_{new} = \begin{cases} 1 & \text{if } num\_new\_grab > 0 \\ 0.5 & \text{otherwise} \end{cases}, \tag{11}$$

$$r_{same} = exp\left[-num\_same\_grab\right], \tag{12}$$

$$r_{grab} = 1.5\left(r_{num\_grab} \times r_{time} \times r_{new} \times r_{same}\right). \tag{13}$$

양팔만 사용하는 것을 방지하기 위해, 양팔에 해당하는 토크의 총합이 클수록 페널티를 부여한다.

$$r_{force} = \frac{1}{2} exp\left[-0.00001 \sum \tau_{arm}^2\right]. \tag{14}$$

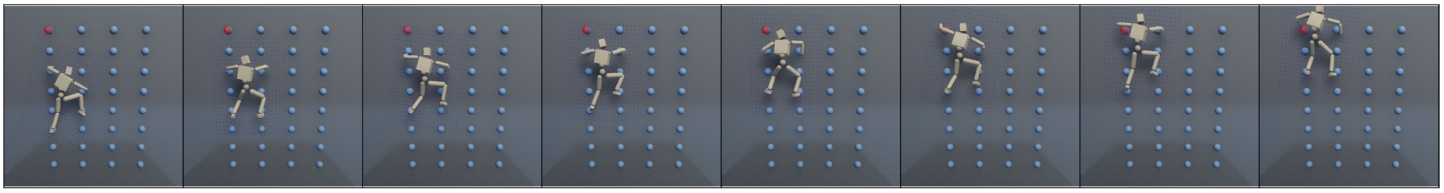
클라이밍 최종 보상 함수는 다음과 같다.

$$r_{climb} = r_{pos} + r_{mov} + r_{speed} + r_{grab} + r_{chest} \times r_{up} + r_{force}. \tag{15}$$

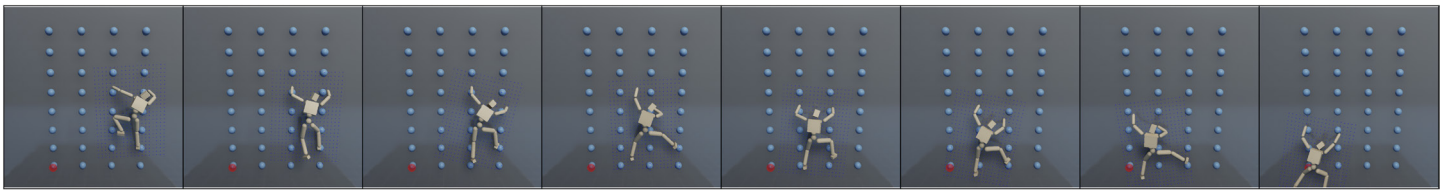
(15)의 수식을 이용해서 학습함으로써, 캐릭터는 물리적으로 그럴싸한 클라이밍 모션을 생성하며 타겟 위치까지 이동할 수 있다.

5 실험 및 평가

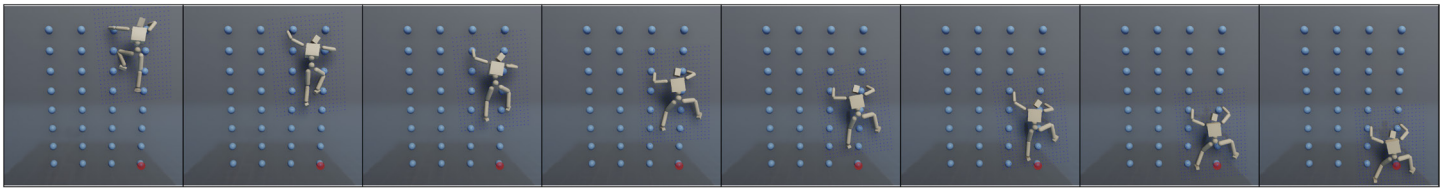
제안하는 방법의 유효성을 평가하기 위해, 모션 품질에 대한 정성 평가와 클라이밍 성공률에 대한 정량 평가를 수행한다. 분류기 학습을 위한 모션 캡처 데이터세트는 매달리기 정책과 클라이밍 정책 각각 다르게 사용한다. 매달리기 정책은 2초 분량의 Mixamo[13] 데이터세트에서 얻은 위로 클라이밍 하는 모션 클립을 사용했고, 클라이밍 정책은 24초 분량의 4개의 모션 클립으로 구성한다. Mixamo[13] 데이터세트에서 위, 아래로 클라이밍 하는 모션 클립 2개, CIMI4D[14] 데이터세트에서 추출한 오른쪽, 왼쪽으로 클라이밍 하는 모션 클립 2개를 사용한다. 강화학습 알



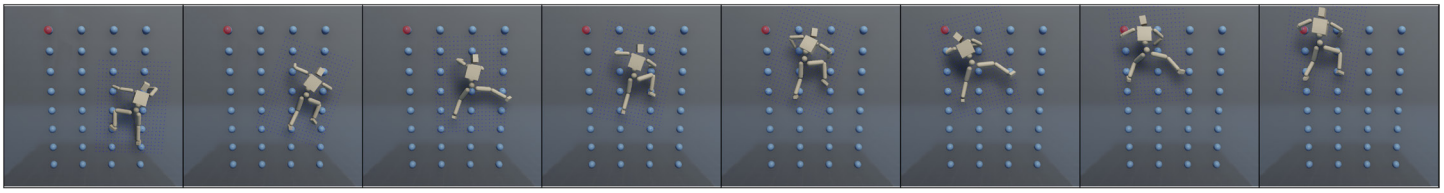
(a) Climb-Up



(b) Climb-Sidestep



(c) Climb-Down



(d) Complex Movement

Figure 4: Trajectories obtained from the learned climbing expert model.

고리즘은 PPO[15]를 사용했고, 분류기, 액터, 크리틱 모두 동일하게 2개의 은닉층 [1024, 512]으로 구성한다. 강화 학습 환경을 위한 시뮬레이터는 Isaac Sim[16] 시뮬레이터를 사용한다. 시뮬레이션 빈도수는 120Hz, 강화 학습 제어 정책의 빈도수는 60Hz 이다. 실험은 한 개의 RTX 4090 GPU을 이용해서 4096개의 환경을 병렬 학습했다.

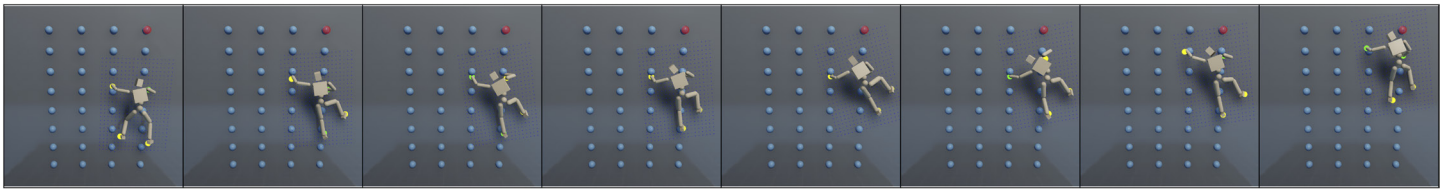
5.1 정성 평가

본 장에서는 두 단계로 구성된 제안하는 방법들의 결과와, 제안하는 방법과 주요 요소들을 제거한 실험 결과들을 정성적으로 비교한 결과를 보인다.

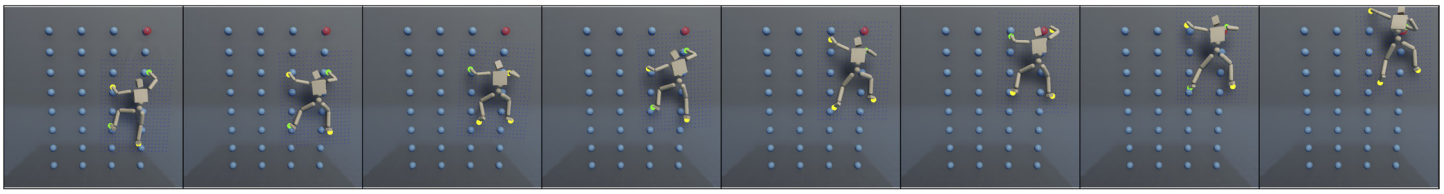
Figure 6은 매달리기 정책의 학습 결과 예시이다. 클라이밍 벽 주변 무작위 위치에서 캐릭터를 떨어뜨렸을 때, 모든 사지를 이용해 다양한 자세로 홀드를 붙잡고 있는 모습을 확인할 수 있다. 실험 환경에서 4개의 홀드를 동시에 붙잡는 상태를 50개 추출하여 초기 상태 데이터세트를 구성했다.

Figure 4는 클라이밍 모델의 학습 결과이다. (a)는 좌측 최상단의 타겟 위치를 향해 위로 올라가는 모습을 보인다. 사지를 적절히 분배하여 올라가는 모습을 확인할 수 있다. (b)는 좌측 최하단의 타겟 위치까지 수평 이동을 수행한다. 홀드를 잡고 있는 왼쪽 팔 위치로 오른팔 위치를 옮기면서, 서로 잡기 상태를 전환하고 재빠르게 다시 왼쪽 팔을 다음 홀드로 이동하며 움직이는 모습을 보인다. 분류기 학습에 사용한 좌, 우 이동 모션 클립에 없는 모션이지만, 제안하는 방법을 사용한 결과 적절한 모션이 합성되었다. 하지만, 이런 역동적인 움직임을 보일 때는 대부분 다리는 잡기를 수행하지 않는 것을 확인할 수 있다. (c)는 우측 최하단의 타겟 위치를 향해서 내려가는 모습을 보인다. 떨어져서 밑에서 잡는 형태가 아니라, 홀드들을 잡고 내려오는 것을 확인할 수 있다. 하지만, (b)와 마찬가지로 다리는 공중에 떠있는 상황이 대부분이다. 올라가기와 달리 굉장히 빠른 속도로 내려오는데, 이 경우에는 타겟 속도를 고정하지 않고 동적으로 변하게 학습시키면 나아질 것이라 추정한다. (d)는 좌측 최상단 타겟을 향해 평행이동과 수직이동이 동반된 복잡한 움직임이다. 단순히 한 방향으로 움직이는 것이 아닌, 복잡한 방향으로도 잘 움직일 수 있다는 것을 확인할 수 있다.

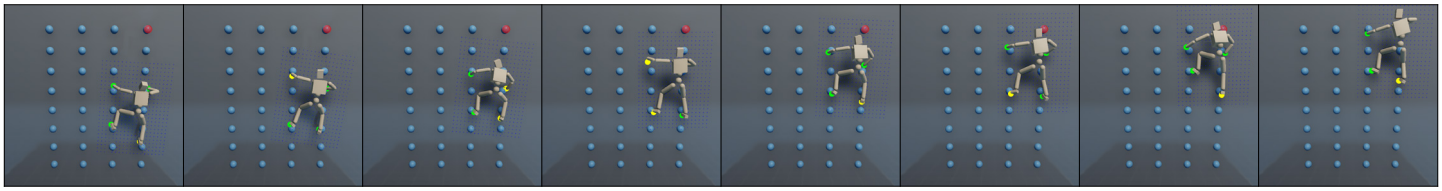
Figure 5는 제안하는 방법의 유효성을 확인할 수 있는 실험 결과이다. 제안하는 방법과 초기 상태 데이터세트 없이 학습한 결과, r_{grab} 없이 학습한 결과를 나타낸다. 캐릭터 사지 말단에 부착된 빛은 잡기 상태를 나타낸다. 초록색은 잡고 있는 상태이고, 노란색은 놓기 상태를 나타낸다. 모두 같은 시작 위치, 자세, 타겟 조건으로 실험하였다. (a)는 초기 상태 데이터세트 없이 클라이밍



(a) No initial states collection



(b) No grab reward



(c) Ours

Figure 5: Ablation Results. The lights attached to the body indicate whether the body is actually anchored. Green : anchored, Yellow : released.

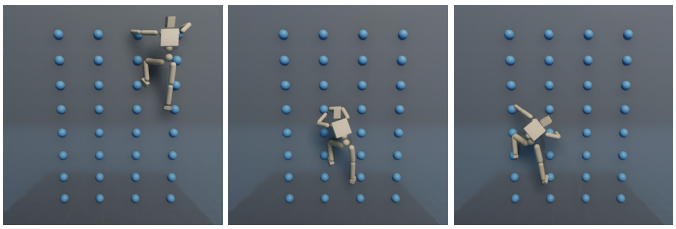


Figure 6: Examples of learning outcomes of the hang expert model.

	Success Rate
Without ISC	0.89
Without grab reward	0.90
Ours	**0.96**

Table 1: Comparison of success rate in our method without initial states collection and without grab reward.

작업을 학습한 결과이다. 양팔만을 교차해 이용해서 올라가고, 진행할수록 자세의 균형이 올바르지 않은 것을 확인할 수 있다. 이를 통해서 초기 상태 데이터세트가 모델에게 동시에 여러 홀드를 잡는 것을 경험시킴으로써, 더 자연스러운 클라이밍 자세를 합성시킬 수 있다는 것을 증명한다. (b)는 r_{grab} 없이 학습한 결과이다. 초기 상태 데이터세트가 있기 때문에 그럴싸한 모션을 보이며 올라가지만, 오른발이 잡기를 하지 않고 공중을 발차는 것을 볼 수 있다. 이를 통해서 r_{grab}가 사지의 잡기 상태를 유도시키는 것을 확인할 수 있다. (c)는 제안하는 방법으로 학습한 결과이다. 제안하는 방법의 실험 결과는 사지 전부를 적절한 순간에 이용하여 올라가는 것을 확인할 수 있다.

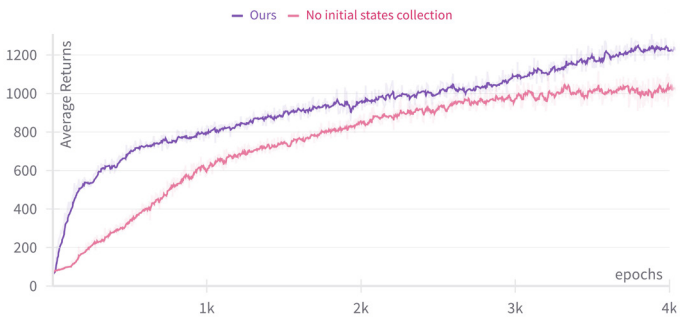


Figure 7: Learning curves comparing performance on climbing task using different methods. We compared ours with no initial states collection.

5.2 정량 평가

본 장에서는 제안하는 방법의 실험 결과와 주요 요소들을 제거한 실험 결과들을 정량적으로 비교한 결과를 보인다.

제안하는 방법이 클라이밍 작업에 성공률을 높인다는 것을 확인하기 위해, 1000 에피소드 동안 승률을 측정하였다. 한 에피소드 동안 무작위로 설정된 타겟 위치와 10cm 이내에 루트가 존재하면 1회 승리한 것으로 판정한다. Table1은 세 방법의 승률을 기록한 표이다. 제안하는 방법의 승률이 0.96로 가장 높은 것을 보인다. 제안하는 방법에 비해서 다리 사용률이 적은 두 비교군은, (14)의 r_{force} 때문에, 극단적으로 다리를 사용하지 않고 양팔의

힘으로만 이동하는 경우는 없었고, 다리가 공중에 떠있어 자세를 잘못 잡아 실패하는 경우가 있는 것을 확인했다.

Figure 7은 제안하는 방법과 초기 상태 데이터세트 없이 학습한 방법의 학습 기간 동안 평균 보상 값을 나타내는 그림이다. 제안하는 방법이 평균 보상이 항상 더 높은 것을 확인할 수 있다. 안정된 자세로 시작하는 제안하는 방법과는 달리, 무작위 위치와 자세로 시작하는 비교군은 초반 학습의 평균 보상이 훨씬 낮은 것을 확인할 수 있다.

6 결론 및 향후 연구

본 논문에서는 2단계로 구성된 강화학습을 이용한 클라이밍 모션 합성 방법론을 제안한다. 매달리기 정책을 통해 유효한 매달리기 상태 데이터 세트를 획득하고, 이를 초기 상태 데이터세트로 사용해서 클라이밍 정책을 학습시킨다. 제안하는 방법이 높은 성공률과 더 자연스러운 자세를 사용하는 것을 증명한다. 하지만, 제안하는 방법은 여전히 많은 한계가 존재한다. 첫 번째로, 클라이밍 벽의 상태(기울기, 홀드 위치)가 변하면 매달리기 모델을 통해 얻은 데이터세트를 사용할 수 없는 단점이 존재한다. 본 논문에서 한 가지 클라이밍 벽을 사용했지만, 여러 클라이밍 벽을 함께 학습함으로써 이러한 한계를 개선할 수 있을 것이라 추정한다. 두 번째로, 보상 함수 수식이 복잡하다. 최신 강화학습 기반 캐릭터 제어 논문들은 보상 함수 수식을 간편화하는 방향으로 연구가 발전하고 있다. 보상 함수 수식을 조정하는 것은 많은 시행착오를 겪어야 하기 때문이다. 미래에는 이러한 한계를 해결하기 위해 모션 캡처 클립에서 잡기 정보를 응용하여 보상 함수 수식 복잡성을 줄이는 것을 목표로 하고 있다. 마지막으로, 제안하는 방법에서 잡기를 구현하기 위해 사용된 제약 조건은, 거리 조건만 충족하면 매달릴 수 있다. 이러한 특성은 사실적인 물리 시뮬레이션 환경과 거리가 멀다는 한계가 있다. 미래에는 이러한 제약 조건을 사용하는 것이 아닌, 실제 캐릭터의 손과 발을 통해 홀드와 상호작용하는 방법에 대한 개발을 계획하고 있다.

감사의 글

이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No. 2021-0-00320 , 실 공간 대상 XR 생성 및 변형/증강 기술 개발)과 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2023-00222776).

References

[1] S. Maheshwari, R. Narain, and R. Hebbalaguppe, "Transfer4d: A framework for frugal motion capture and deformation transfer," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12 836–12 846, 2023.

[2] X. Yi, Y. Zhou, and F. Xu, "Transpose: Real-time 3d human translation and pose estimation with six inertial sensors," *ACM Transactions on Graphics*, vol. 40, no. 4, 2021.

[3] X. B. Peng, P. Abbeel, S. Levine, and M. van de Panne, "Deepmimic: Example-guided deep reinforcement learning of physics-based character skills," *ACM Trans. Graph.*, vol. 37, no. 4, pp. 143:1–143:14, July 2018.

[4] X. B. Peng, Z. Ma, P. Abbeel, S. Levine, and A. Kanazawa, "Amp: Adversarial motion priors for stylized physics-based character control," *ACM Trans. Graph.*, vol. 40, no. 4, July 2021.

[5] X. B. Peng, Y. Guo, L. Halper, S. Levine, and S. Fidler, "Ase: Large-scale reusable adversarial skill embeddings for physically simulated characters," *ACM Trans. Graph.*, vol. 41, no. 4, July 2022.

[6] J. Ho and S. Ermon, "Generative adversarial imitation learning," in *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[7] C. Tessler, Y. Kasten, Y. Guo, S. Mannor, G. Chechik, and X. B. Peng, "Calm: Conditional adversarial latent models for directable virtual characters," in *ACM SIGGRAPH 2023 Conference Proceedings*. New York, NY, USA: Association for Computing Machinery, 2023.

[8] Z. Dou, X. Chen, Q. Fan, T. Komura, and W. Wang, "C · ase: Learning conditional adversarial skill embeddings for physics-based characters," *arXiv preprint arXiv:2309.11351*, 2023.

[9] Q. Zhu, H. Zhang, M. Lan, and L. Han, "Neural categorical priors for physics-based character control," *ACM Trans. Graph.*, vol. 42, no. 6, dec 2023. [Online]. Available: https://doi.org/10.1145/3618397

[10] K. Naderi, J. Rajamäki, and P. Hämäläinen, "Discovering and synthesizing humanoid climbing movements," *ACM Trans. Graph.*, vol. 36, no. 4, jul 2017.

[11] K. Naderi, A. Babadi, and P. Hämäläinen, "Learning physically based humanoid climbing movements," *Computer Graphics Forum*, vol. 37, no. 8, pp. 69–80, 2018.

[12] K. Naderi, A. Babadi, S. Roohi, and P. Hämäläinen, “A reinforcement learning approach to synthesizing climbing movements,” in *2019 IEEE Conference on Games (CoG)*, 2019, pp. 1–7.

[13] Adobe, “mixamo,” *https://www.mixamo.com*, 2020.

[14] M. Yan, X. Wang, Y. Dai, S. Shen, C. Wen, L. Xu, Y. Ma, and C. Wang, “Cimi4d: A large multimodal climbing motion dataset under human-scene interactions,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 12 977–12 988.

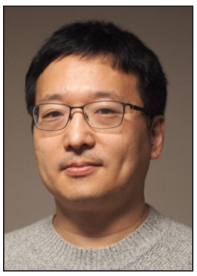
[15] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” *CoRR*, vol. abs/1707.06347, 2017. [Online]. Available: http://arxiv.org/abs/1707.06347

[16] V. Makoviychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa, and G. State, “Isaac gym: High performance gpu-based physics simulation for robot learning,” *CoRR*, vol. abs/2108.10470, 2021. [Online]. Available: https://arxiv.org/abs/2108.10470

〈저자소개〉

강경원

- 2016–2020 중부대학교 정보보호학 2학년 수료(한양대학교 ERICA 편입)
- 2020–2022 한양대학교(ERICA) 소프트웨어학 학사
- 2022–현재 한양대학교 컴퓨터소프트웨어학 석사과정
- https://orcid.org/0009-0004-7328-1760



권태수

- 1996–2000 서울대학교 전기컴퓨터공학부 학사
- 2000–2002 서울대학교 전기컴퓨터공학부 석사
- 2002–2007 한국과학기술원 전산학전공 박사
- https://orcid.org/0000-0002-9253-2156